ISRN TELCORDIA--2011-05+PR-0GARAU
Prepared for: Office of Naval Research

Fast Multiscale Algorithms for Information Representation and Fusion

Technical Progress Report No. 5

Devasis Bassu, Principal Investigator

Contract: N00014-10-C-0176

Telcordia Technologies

One Telcordia Drive

Piscataway, NJ 08854-4157

October 2011

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **OCT 2011**	2. REPORT TYPE	3. DATES COVERED
4. TITLE AND SUBTITLE **Fast Multiscale Algorithms For Information Representation And Fusion**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Telcordia Technologies,One Telcordia Drive,Piscataway,NJ,08854**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**14**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

1 Abstract

In the fifth quarter of the work effort, we focused on a) conducting experiments on real-world data sets using the developed algorithms, b) design of the Multiscale Singular Value Decomposition (SVD) algorithm, c) preliminary design of the Multiscale Heat-Kernel Coordinates algorithm and d) fine tuning and bug fixes for the randomized SVD and ANN algorithms. This report documents the line-of-sight experiments using radio-frequency (RF) data.

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and implement the multiscale SVD and heat-kernel coordinates algorithms. No problems are currently anticipated.

Table of Contents

2 Summary

In this quarter, we conducted line-of-sight experiments on real-world RF data sets in addition to designing the new multiscale SVD and heat-kernel coordinates algorithms.

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and implement the multiscale SVD and heat-kernel coordinates algorithms. No problems are currently anticipated.

3 Introduction

The primary project effort over the last quarter focused on experimentation with a real-world line of sight RF dataset along with bug fixes and fine-tuning the randomized SVD and ANN algorithms [1][2]. We continued work on the design of the new multiscale Singular Value Decomposition and Heat-Kernel coordinates algorithms [3][4]. A detailed write-up of the line of sight experiments is provided in Section 4.

4 Methods, Assumptions and Procedures

4.1 Line of Sight Experiments

The following describes the line of sight experiments conducted in this quarter using the developed algorithms.

The objective here is to answer if one can determine if a received signal consists of signal+multipath, or direct signal only using sampled RF signal measurements. Measurements are performed by placing a signal source in a known location and driving a route with several types of multipath conditions. This knowledge is important in geolocation applications where knowing whether a received signal is line-of-sight (or not) is necessary for other algorithms to work.

This data set used in this experiment was collected on-site at the Telcordia Navesink campus at Red Bank, NJ before the start of this project.

4.1.1 Description of the Dataset

Two sets of DRS 9144(receiver)/9475(digitizer) pairs were used along with GPS receivers/loggers in addition to signal generators and antennas to generate and collect the RF data. Each different multipath condition (full/partial/zero line-of-sight) along the route was time-stamped and recorded. The dataset comprises three repeated trials of GPS time-stamped raw baseband I+Q measurements obtained from driving a truck with the receivers along a pre-determined path around the Telcordia Navesink campus. Each I/Q values are 16-bit integers. Each time-stamped record contains 4 I/Q value pairs. The sampling interval is 4 microseconds.

Trial 1

Duration: 8+ mins

Number of records: 122,591,820

Size: 1,961,469,120 bytes

Trial 2

Duration: 7+ mins

Number of records: 112,057,920

Size: 1,792,926,720 bytes

Trial 3

Duration: 8+ mins

Number of records: 121,082,910

Size: 1,937,326,560 bytes

Additional metadata including frequency, AGC backoff, RF attenuation is available. The GPS data is provided in a separate file in the GPX format with associated reading timestamps. The location of the transmitting antenna along with frequency, field strength and power parameters are also available.

4.1.2 Data Analysis Methods and Procedures

Preprocessing

Each individual I/Q data stream is converted into a power (dB) data stream using

$$s(t) = \sum_{j=t}^{t+\delta-1} 10 * log10\left(I_j^2 + Q_j^2\right) + signal\ attenuation - 148$$

where δ is a suitable time window.

Data Matrix

Power data streams for each of the four available channels are combined with a suitable lag l to form the data matrix P that is used for subsequent analyses. Each row (data point) is defined as

$$P_{t,*} = [s_t^1, s_{t+1}^1, \dots, s_{t+l-1}^1, s_t^2, \dots, s_{t-l+1}^2, s_t^3, \dots, s_{t+l-1}^3, s_t^4, \dots, s_{t+l-1}^4]$$

where s_t^c is the power reading at time t for channel c.

The central idea to both the analysis techniques outlined below is to obtain a reduced-dimensional representation that effectively captures the system dynamics using the matrix P. Each data point $P_{t,*}$ is transformed to its corresponding point in the lower-dimensional space. System states, or operations regimes, are defined on the lower-dimensional space. This allows one to detect states (or observe anomalous behavior) for an online streaming setup. In this experiment, we want the states uncovered by the analyses to reflect various line of sight conditions.

SVD Analysis

Here, we perform a Singular Value Decomposition (SVD) of the matrix P to obtain the singular values and singular vector corresponding to the space spanned by the row vectors $P_{t,*}$. The singular values provide insight for selecting the reduced dimensionality of the target space. The corresponding singular vectors provide the projection into this reduced dimensional space.

Diffusion Map Analysis

First, we construct the normalized graph Laplacian matrix M for the points $P_{t,*}$. Next, we proceed with SVD analysis of M as outlined above (in this case, we get the eigenvalues and eigenvectors) to obtain the reduced dimensional representation.

The experimental results are described in Section 5.

4.2 Deliverables / Milestones

Date	Deliverables / Milestones	Status
Oct 2010	Progress report for period 1, 1st quarter	✓
Jan 2011	Progress report for period 1, 2nd quarter / complete randomized matrix decompositions task	✓
Apr 2011	Progress report for period 1, 3rd quarter / complete approximate nearest neighbors task	✓
Jul 2011	Progress report for period 1, 4th quarter / complete experiments – part 1	✓
Oct 2011	Progress report for period 2, 1st quarter	✓
Jan 2012	Progress report for period 2, 2nd quarter / complete multiscale SVD task	
Apr 2012	Progress report for period 2, 3rd quarter	
Jul 2012	Progress report for period 2, 4th quarter / complete experiments – part 2	
Oct 2012	Progress report for period 3, 1st quarter	
Jan 2013	Progress report for period 3, 2nd quarter / complete multiscale Heat Kernel task	
Apr 2013	Progress report for period 3, 3rd quarter	
Jul 2013	Final project report + software + documentation on CDROM / complete experiments – part 3	

The next section contains experimental results for the line of sight datasets.

5 Results and Discussion

We present selected results from various experiments conducted on the line of sight datasets.

5.1 Experimental Setup and Ground Truth

The raw data is sampled every 4 microseconds. For each channel, two streams of power data were computed with 1 millisecond and 1 second intervals. We tested with a variety of lag sizes ranging from 5 to 30. While there are no radical differences between the choices for the lag size, details tend to disappear as the lag size increases due to averaging effects.

Figure 1 shows an ariel view of the Telcordia Navesink campus along with the location of the transmitting antenna. Knowledge of the environment and GPS readings provided us with ground truth. The red star shows the location of the transmitter.

Figure 1: Telcordia Navesink Campus

The raw data was processed and analyzed using techniques described in Section 4. Selected analysis results are presented below.

5.2 SVD Analysis

Figure 2 shows the singular values for the computed SVD on the data matrix (described earlier in Section 4) with a 5 second lag. It indicates that most of the information is effectively captured in just the first dimension; the second dimension may provide some finer details.

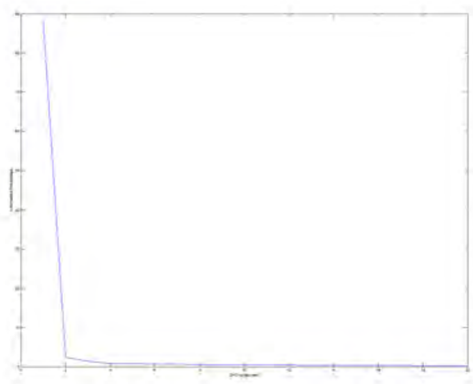

Figure 2: Singular values for power data matrix

Figures 3 and 4 depict the path driven using colored markers. The colors correspond to the projection values on the first and second singular vectors respectively. The color scheme is the JET color map which ranges from dark blue to dark red with green in the middle.

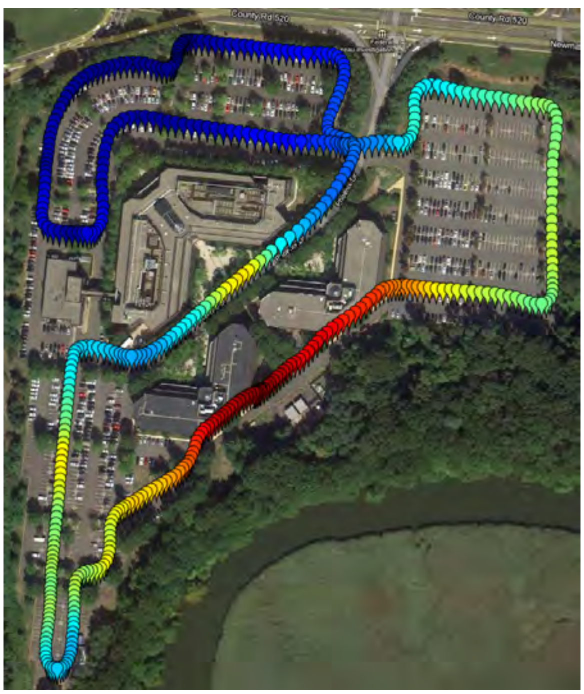

Figure 3: Projection using 1st singular vector

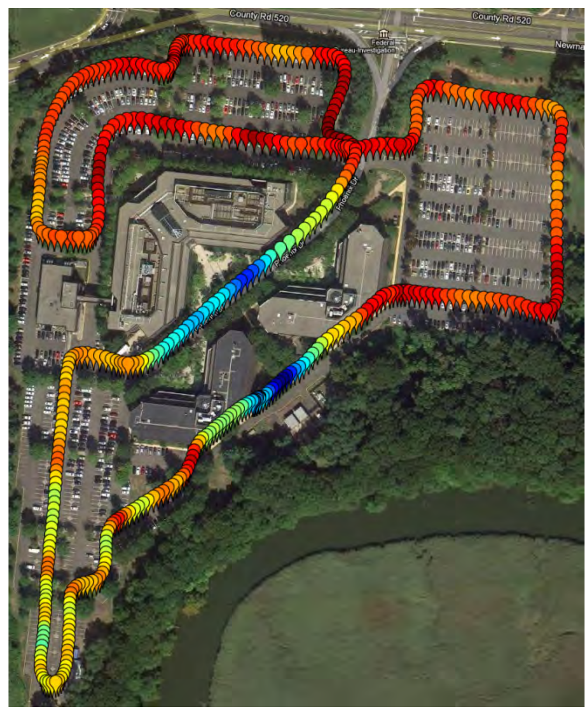

Figure 4: Projection using 2nd singular vector

Figure 3 effectively captures the various line of sight conditions. Dark red indicates the area right next to the transmitter (see Figure 1) which provides full line of sight. In contrast, dark blue (top-left loop) indicates no line of sight. Other intermediate color indicate partial line of sight. The differences between the bottom-left and top-right segments, which are both partial line of sight in Figure 3, are highlighted if we look at Figure 4 which indicates that the bottom-right segment has slightly better line of sight conditions than the top-right segment. This is in agreement with the ground truth as specified by a domain expert. In summary, a 2-dimensional space may be sufficient to determine various line of sight conditions from the RF signals (a 1-dimensional space may be suitable where finer distinctions are not required).

5.3 Diffusion Map Analysis

Figure 5 shows the eigenvalues for the graph Laplacian data matrix (described earlier in Section 4) with a 5 second lag. As in the case of the SVD analysis, a 2-dimensional space provides adequate representation. Further, the line of sight conditions differ only by minor amounts between the SVD and the diffusion map analyses suggesting an approximately linear structure on the points in the power data matrix.

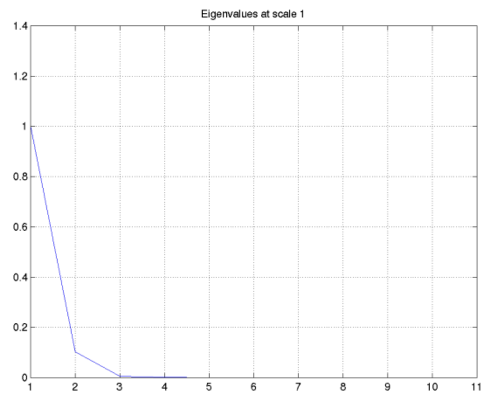

Figure 5: Eigenvalues for graph Laplacian of power data matrix

5.4 Discussion

The analysis results show an effective characterization of the various line of sight conditions in the computed low-dimensional spaces. Additional experimentation was conducted to account for loss of signal based on distance from the transmission point by making suitable adjustments to the power measurements. The results were marginally different to be conclusive.

All the trials in the current dataset involve driving along the same route. Data from additional routes would be helpful in testing the generalization of the low-dimensional spaces obtained as part of training.

6 Conclusions

The project is on track with experimentation on real-world data sets using the developed algorithms as well as the continued design of the multiscale SVD and heat-kernel coordinates algorithms. We will continue algorithm design and experimenting on the selected real-world data sets using the developed algorithms in the next quarter.

No problems are currently anticipated.

7 References

[1] V. Rokhlin, M. Tygert, *A fast Randomized Algorithm for the Overdetermined Linear Least Squares Regression*, PNAS, vol. 105, No. 36, pages 13212-13217, 2008.

[2] P. Jones, A. Osipov, V. Rokhlin, *A Randomized Approximate Nearest Neighbors Algorithm*, Research Report YALEU/DCS/RR-1434, Yale University, September 14, 2010

[3] G. Lerman, *Quanitfying curvelike structures of measures by using Jones quantities*, C.P.A.M., vol. 56, issue 8, pages 1294-1365.

[4] P.W. Jones, M. Maggioni, R. Schul, Manifold parameterizations by eigenfunctions of the Laplacian and heat kernels, PNAS, vol. 105, no. 6, pages 1803.